Studying Interdependencies in Music Performance: An Interactive Tool

Tilo Hähnel, Axel Berndt
Otto-von-Guericke Universität Magdeburg
Universitätsplatz 2
39106 Magdeburg
{tilo,aberndt}@isg.cs.uni-magdeburg.de

ABSTRACT

Musicians tend to model different performance parameters intuitively and listeners seem to perceive them, to a certain degree, unconsciously. This is a problem for the development of synthetic performance models, for they are built upon detailed assumptions of several parameters like timing, loudness, and duration—and of interdependencies as well. This paper describes an interactive performance synthesis tool, which allows to analyse listener's preferences of multiple performance features. Using the tool in a study of eighth notes inégalité, a relationship between timing and loudness was found.

Keywords

Synthetic Performance, Notes Inégales, Timing, Articulation, Duration, Loudness, Dynamics

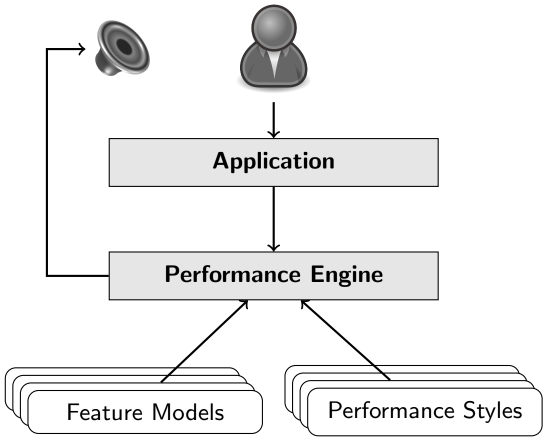

Figure 1: The technical setup of the study.

1. INTRODUCTION

Up to now, timing seemed to be a core feature of expression in music. Despite large phrase arches and ritardandi, it is the temporal shape of small figures, which is crucial for the impression of liveliness and expressiveness. Gabrielsson analyzed ratios of beats or frequent notes at the sub-beat level [6]. He discovered that notes of equal value are performed with different ratios—even in several rhythms of a broad stylistic range, reaching from Swedish folk songs to the Viennese Waltz. Similar phenomena were also discovered in more remote cultures, as Gerischer demonstrated [7], or in diverse Western music styles, which were analysed, for instance, by Langner [11].

The knowledge of an unequal shaping of notes has been known in theory and practise for centuries. A prominent instance is the playing of "notes inégales" in French Baroque, which is discussed in detail by Hefling [8]. Simply put, playing eighths notes inégal means that a note on the beat lasts a little longer than its actual value. The time added on this note is subtracted from the succeeding eighth note between the beat. Inégalité therefore describes a particular phenomenon of timing. However, referring to original sources like Quantz' treatise [13], inégalité has always included aspects of loudness and articulation. Inégalité therefore could be seen as an emphasis of a note, which is prominently achieved by lengthening the first eighth note (timing), but also by playing it louder (loudness) or articulating it differently, e.g. by manipulating the duration of a note. An analysis of inégalité should therefore take into account the complex interplay of at least timing, loudness, and tone duration.

To focus on interdependencies between these three performance features, we developed a tool, with which listeners could adjust these parameters interactively. (see Section 2 and 3). The main idea is to extend the common "Analysis-by-Synthesis" approach that was outlined in detail by Bengtsson and Gabrielsson [1] and reaches back to Seashore and his colleagues, like Metfessel [14, 12]. Normally, listeners are asked to judge several stimuli, which comprised synthetic performances that differed in some particular performance parameters. The listener's judgements then indicate which parameters are most appropriate.

This approach had to be modified, for the total amount of stimuli depends on the number of grades a parameters is subdivided into and increases to the power of parameters used. This would mean that a combination of three parameters of 21 grades each would result in $21^3 = 9261$ stimuli, an amount impossible to be judged by listeners. The task was to provide all stimuli but at the same time reduce them to a minimum for each participant. This was done by letting the participants manipulate these three parameters independently and interactively until they reached the parameter combination they preferred.

The following Section includes a description of the tool. Section 3 describes the method of the inégalité study. Finally, a general discussion closes the paper in Section 5.

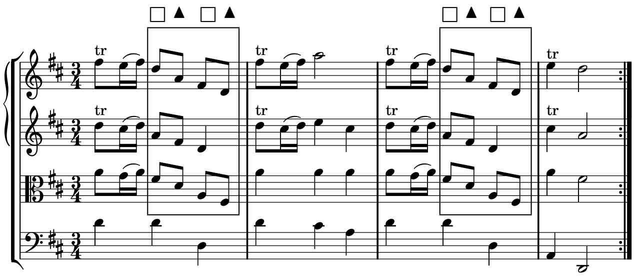

Figure 2: First bars of the Polonoise taken from the overture in d-major "La Gaillarde" TWV 55:D13 for strings & b.c., composed by G. P. Telemann.

2. TECHNICAL SETUP

We have developed a performance rendering engine that allows to create expressive music performances based on a high-level description language (XML-based). Different performances can be described and are stored as *performance styles*. These are rendered into expressive MIDI sequences. The performance engine furthermore allows to interactively change over to another performance style while the music plays. It automatically creates musically plausible seamless transitions by simulating the reactions of musicians with an agent-based approach [2]. This allows any application to interactively control the music performance just like in the case of the study described in this paper. The participants had to explore a domain of possibilities and choose the location that matches their preferences. The setup is shown in Figure 1.

Formally, expressive performance can be regarded as a series of transformations that is applied to a musical composition. We distinguish three categories:

Timing defines the mapping from symbolic time, which is where the composition is defined in, onto physical time (usually in milliseconds). It combines several layers of timing features, i.e., tempo (macro timing), rubato (self-compensating micro deviations), asynchrony (between parts in a polyphonic setting), and random imprecision. [3, 4]

Dynamics sets the loudness of each musical event. It incorporates a macro layer that defines the basic loudness function of the musical piece, and two micro layers that introduce fine deviations to the basic loudness. Micro dynamics features are metrical accentuations and articulations (only the loudness aspect of the latter). [5]

Articulation deals with the way each musical event (note) is played. Articulation instructions affect the loudness, duration and timbre of the notes to be articulated. [9, 10]

We have developed mathematical models to describe, analyze, and synthesize these features. Their flexible parameterization allows for a huge bandwidth of characteristics, including those that could be observed with human musicians, and even atypical ones that are more of theoretical interest. Most models do even allow to define new instances, e.g., new articulations, accentuation schemes, and rubato patterns. This flexibility and full control over any performance detail, even the most subtle, makes our performance engine a very useful tool especially for musicological and psychological experiments and studies.

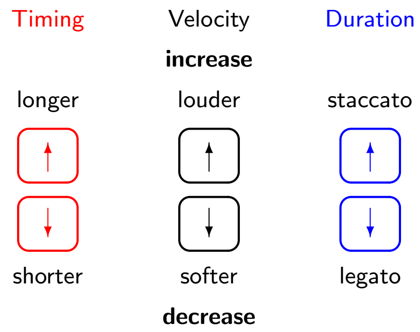

Figure 3: Instructions for the modification of eighth notes occurring on the beat.

3. METHODOLOGY

Timing, loudness, and duration were tested by the use of a four bar phrase taken from a Polonoise by G. P. Telemann (see Figure 2).

Two tests were carried out, first a *separate parameter test* and then a *combined parameter test*. In the first test the parameters were analysed separately. The participants were asked to modify the performance of eighth notes labelled in a score (as shown in Figure 2) of the stimulus, which was presented in a loop. The eighth notes on the beat (□ in Figure 2) were emphasized by pressing an Arrow-Up key. A decrease of this emphasis or even an emphasis of the eighths between the beats (▲ in Figure 2) was set by pressing the Arrow-Down key. At the limit of the parameter spectrum a beep signalled that no further modification was possible in that direction.

The first test consisted of two tasks: First, the participants were asked to identify the parameter they set. Then they had to tune the relation of the eighths as they supposed to sound best and confirm with the Enter-key. All participants only saw the score including the squares and triangles as shown in Figure 2 and therefore depended completely on an auditive feedback.

In the second test the same stimulus was presented. The participants now had the opportunity to tune all three parameters independently at the same time. As Figure 3 shows, all parameters were set by an array of six control-keys above the arrow keys, which were labelled with up and down arrows in red, black and blue.

3.1 Feature Manipulation

Our performance engine implements two output modes, a standard MIDI mode and further functionalities for the output over the software sampler *Vienna Instruments* that utilizes some its advanced possibilities in controlling certain tone parameters. In this study, however, we applied the software *VSampler 3* that runs smoothly on laptops, allowing for more flexibility and mobility.

Via the study application the participants of the study had control over certain parameters that influence the timing (rubato), dynamics (metrical accentuation), and articulation (tone duration). The interface is shown schematically in figure 3. The parameter space was discretized into 21 steps (controller states) reaching from -10 (minimal setting) to 10 (maximal setting).

Rubato: Rubato is a timing distortion that is compensated within a certain timeframe. Hence, the basic tempo remains unchanged at 120 bpm. In this study the rubato frame was of the length of a quarter note and repeatedly applied over the whole musical piece. For each quarter note a swing-like distortion could be cre-

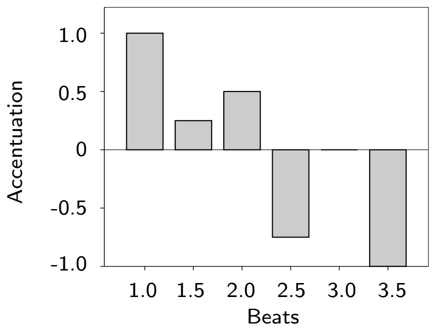

Figure 4: The accentuation scheme could interactively be scaled.

ated. This distortion is modelled by a potential function in the unity square. The timing relation between first and second eighth is:

$$\Big((0.5^i)\,0.98 : (1 - 0.5^i)\,0.98\Big) \quad | \quad i \in [0.4, 1.6]$$

The timing controls enabled the probands to set the parameter i. Values between $10 \rightarrow (i = 0.4)$ and $-10 \rightarrow (i = 1.6)$ were linearly interpolated.

Metrical Accentuation: On the level of micro-dynamics an accentuation scheme was defined (see figure 4) and applied to each measure. The participants were able to set its intensity, that is the loudness scale of the scheme, ranging -60 up to 60 MIDI velocity units. The possible settings reached from 10 to -10, whereas 10 created a pianissimo for the softest and a fortissimo for the strongest accentuations (a range of 107 MIDI velocity units) and -10 caused a likewise pronounced but inverted accentuation scheme.

Articulation: The probands could set the durations of the notes from legato to a very short staccato. This cause either a very cantabile or a rhythmically pronounced performance. However, the behaviour of articulation settings was even more complex. The duration of each second eighth note under a quarter beat decreased faster than the first so that the emphatic relation between both shifts accordingly. Three sampling points were defined (see the following table) and linearly interpolated.

controller state	Duration of 1st eighth	Duration of 2nd eighth
10	0.35	0.2
0	0.7	0.4
-10	1.0	1.0

In the *separate parameter test* the controllers were randomly initialized at the extremes, i.e. -10 or 10, whereas, in the *combined parameter test* the initialization was random over the full parameter space. Positive controller settings created plausible features that could also be observed within human performances. With regard to metrical accentuation and rubato, the negative settings created implausible performances.

All interactions were logged, as well as the test duration and the initial and final controller settings that the participants approved. This allows insights into how the users explored the search space and how long they listened to certain settings until they made a decision and interaction.

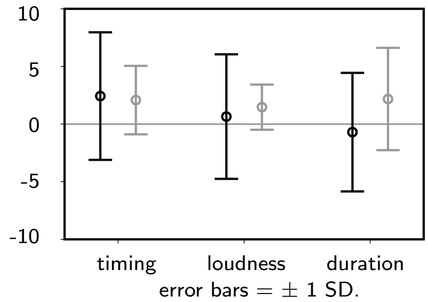

Figure 5: Results of separate tuning. Values tuned by participants identifying parameters correctly (gray bars) compared to others (=black bars).

3.2 Participants

The participants comprised 36 western socialized adults, including 21 women and 15 men, 10 professional musicians specialized in Baroque performance, 16 with a degree in music, musicology, music pedagogy, church music or similar, and 20 playing an instrument for more than ten years.

4. RESULTS

The answers given in the identification task were collected and manually classified as correct, ambivalent, or incorrect. A correct answer had to be unambiguous. This was a problem regarding the differentiation between timing and duration, for the term “length” is ambivalent.

Unfortunately, the distribution of controller values from the second task was not normal. A possible reason was that obviously some participants did not use a spectrum large enough to get an impression of the possibilities they had (it was also hardly possible to identify the parameter if it had not been modified to a certain degree). Therefore, data were excluded when the modification range during the test was below seven (regardless of the final value). This was the case for two samples—the remaining samples showed a normal distribution.

No influence of expertise on the parameter identification or the adjustment was found, which might be caused by the small amount of professional musicians. However, the label “expert” was restricted to the individual musical background. But those identifying the parameters correctly or ambivalent differed to the remaining in the parameter adjustment: the variances of timing and loudness significantly decreased. As it turned out, the standard deviation sd for participants identifying loudness correctly or ambivalent was $sd = 1.69$ compared to $sd = 5.4$ for those who did not identify what they had modified (with a significance of $p = 0.001$ in a Levine test). For timing these differences were smaller ($sd = 2.95$ in the correct group and $sd = 5.53$ in the incorrect group) but still significant (with $p = 0.026$ in a Levine test). The differences are plotted in Figure 5.

Generally, the difficulties the participants had were more pronounced than expected. Hence, 11 participants did not take part in the second test. The remaining 25 participants included all professional musicians. 19 participants of the second test played an instrument for more than ten years and 15 had a degree in a music related subject.

A correlation analysis uncovered a highly significant negative correlation between timing and loudness (with $r = -0.653$ and $p = 0.001$). The plot in Figure 6 shows every sample pair of timing and loudness values, and a regression line to illustrate the correlation. In contrast, duration was correlated neither with timing nor with loudness. Neither

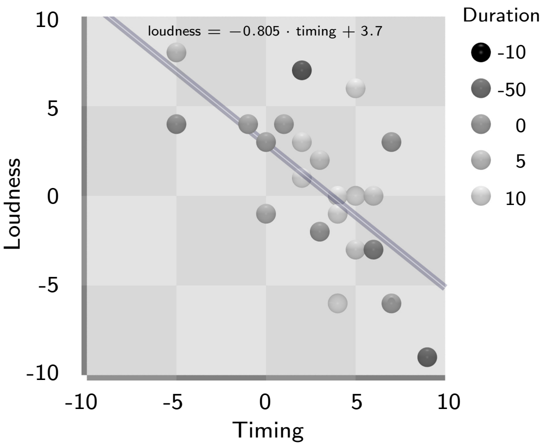

Figure 6: The regression line demonstrates the negative correlation between timing and loudness values.

a t-test nor a Levine-test could detect a difference between the separate and combined modulation of any parameter.

5. GENERAL DISCUSSION

The tool described in the previous Sections allows listeners to find a preferred stimulus out of a large number of stimuli. Generally, the tool can be used for diverse designs of studies with an Analysis-by-Synthesis character. The advantage is that multiple performance features as well as their interdependencies can be tested. In the present study the participants easily understood the tasks and handling of the study application, which is important for listeners unacquainted with alternative interfaces and hardware. Different research questions can nevertheless require different interfaces, which is also not a problem, for the user interface and the performance engine are separate modules. Participants could use sliders instead, move around in a room or produce physical gestures, which can trigger the application. The latter in particular offers interesting perspectives for therapy and rehabilitation, since the setup of stimuli can prompt participants to move in a particular way and avoid other movements. The feature models and performance styles are freely editable and exchangeable. This is particularly interesting because a large amount of stimuli allows a feeling of continuous stimulus modification. This also induces an impression of the direction of change, which the user can consciously track or reverse.

However, the study described focused on interdependencies of timing, loudness, and duration with respect to “inégalité” in eighths notes performances. The results can lead to the suggestion that multiple means for emphasis can be cumulative: an intensive use of a single performance parameter causes the same emphasis—and therefore the same impression of inégalité—as a slight use of multiple performance parameters. To avoid an overemphasis, the musician or music producer must balance out these parameters, which leads to a compensation effect.

The difficulties occurred in the study of “inégalité” uncovered a problem more relevant to a rather musicological discussion. Many participants could not decompose the parameters that are manipulated to achieve inégalité, even if they recognized that something was changing and intended as well as achieved an inégalité while exploring the range of stimuli.

Acknowledgments

We thank all our participants from the department of early music of the UdK in Berlin, the “Zentrum für Telemann-Pflege und -Forschung” in Magdeburg, the Magdeburgian Music school, the Orchestra of the “Magdeburger Musikfreunde”, the “Otto-von-Guericke-Universität Magdeburg” as well as those who do not belong to any of the above mentioned institutions.

6. REFERENCES

[1] I. Bengtsson and A. Gabrielsson. Rhythm research in Uppsala. *Music Room and Acoustics*, pages 19–56, 1977.

[2] A. Berndt. Decentralizing Music, Its Performance, and Processing. In M. Schedel and D. Weymouth, editors, *Proc. of the Int. Computer Music Conf. (ICMC)*, pages 381–388, New York, USA, June 2010. International Computer Music Association, Stony Brook University.

[3] A. Berndt. Musical Timing Curves. In *Proc. of the Int. Computer Music Conf. (ICMC)*, Huddersfield, UK, Aug. 2011. International Computer Music Association, University of Huddersfield. in review.

[4] A. Berndt and T. Hähnel. Expressive Musical Timing. In *Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion*, pages 9–16, Glasgow, Scotland, Sept. 2009. Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå.

[5] A. Berndt and T. Hähnel. Modelling Musical Dynamics. In *Audio Mostly 2010: 5th Conf. on Interaction with Sound—Sound and Design*, pages 134–141, Piteå, Sweden, Sept. 2010. Interactive Institute/Sonic Studio Piteå, ACM.

[6] A. Gabrielsson. The Performance of Music. In D. Deutsch, editor, *The Psychology of Music*, pages 501–602. Academic Press/Elsevier, San Diego, 1999.

[7] C. Gerischer. *O singue baiano - Mikrorhythmische Phänomene in baianischer Perkussion*. Peter Lang, Frankfurt a.M., 2003.

[8] S. E. Hefling. *Rhythmic Alteration in Seventeenth- and Eighteenth-Century Music, Notes Inégales and Overdotting*. Schirmer Books, New York, 1993.

[9] T. Hähnel. From Mozart to MIDI: A Rule System for Expressive Articulation. In *Proceedings of New Interfaces for Musical Expression (NIME2010)*, pages 72–75, Sydney, Australia, June 2010. University of Technology Sydney.

[10] T. Hähnel and A. Berndt. Expressive Articulation for Synthetic Music Performances. In *Proceedings of New Interfaces for Musical Expression (NIME2010)*, pages 277–282, Sydney, Australia, June 2010. University of Technology Sydney.

[11] J. Langner. *Musikalische Rhythmen und Oszillation – Eine theoretische und empirische Erkundung*. Schriften zur Musikpsychologie und Musikästhetik - Band 13. Peter Lang, Frankfurt a.M., 2002.

[12] M. Metfessel. Sonance as a form of tonal fusion. *Psychological Review*, 33(6):459–466, 1926.

[13] J. J. Quantz. *Versuch einer Anweisung die Flöte traversière zu spielen*. Bärenreiter, Berlin, 1752. Faksimile-reprint (1997).

[14] H. Seashore. The Hearing of the Pitch and Intensity in Vibrato. In C. E. Seashore, editor, *The Vibrato*, volume I of *Studies in the Psychology of Music*, pages 213–235. University Press, Iowa, 1932.